Reviewer comments in black
Response in blue
Text in the revised MS in italic

Reviewer #1
This is the second time I review the manuscript by Kohonen et al. I believe the major points have been adequately addressed by the authors. The overall flow and clarity of the manuscript have been improved substantially. I have attached a few minor and technical comments below.
We thank the reviewer for the further comments and suggestions to improve the manuscript.

L14: I believe a comma is advisable before "leading to ...".
Corrected as suggested.

L17: "can be estimated from simple meteorological measurements or the literature" - And leaf and soil hydraulic parameters?
The parameters referred to in this sentence include all the model parameters in Table 1 (meteorological, environmental, leaf photosynthetic, xylem hydraulic, soil hydraulic, …). As this is the abstract, there is no room to go into any further detail about the specific nature of the parameters. Thus, we have simplified the sentence to "*can be estimated from simple measurements or obtained from the literature*".

L53: "proxies for CO2 uptake" -> "proxies for photosynthetic CO2 uptake"
Corrected as suggested

L121: "the mean of the air temperature at 18 m height" - On line 103, you mentioned that air temperature was measured at 16.8 m.
Thank you for noticing this, it was corrected to 16.8m, where the temperature measurements indeed took place.

L200, L251--252: Add units for X and Y.
Added units for X and Y: [X] = MPa m^2 s mol^{-1}; [Y] = mol m^{-2} s^{-1}

L230--233: The absence of a statistically significant difference on the monthly scale feels perplexing (Figs. B4 and B5). Why do GPP_LRU and GPP_CAP differ substantially on 30-min and daily scales but not on the monthly scale? Is there any difference in monthly aggregation or gap-filling? It would be helpful to give a brief explanation as to why this is the case.
The monthly values are means from the daily values, which are calculated from gap-filled 30min flux data so that at least 50% of fluxes need to be measured each day (meaning at least 24 datapoints out of total 48). Averaging reduces scatter, and thus difference to the reference GPP. There is no difference in gap-filling between the time scales, as gap-filling is done to 30min fluxes only, from which the daily and monthly averages then yield. However, there is an increased uncertainty in the ANOVA test on monthly scale, as the sample size is reduced from 23994 (30min flux values) to 30 (monthly fluxes).

L239: "However, there is less scatter ..." - A reference to Fig. 4d,g may be included here.
Figure reference added.

L255: "as the difference between fitted LRU_CAP and literature-based LRU_CAP (statistical significance tested with Student's t-test, $p<0.01$) was not large ..." - I find this statement unintuitive. The parameter Y determines the light dependence of LRU, and if it is increased by a factor of 3.3, the term sqrt(1 + Y / PAR) would increase by a factor of 1.8 when PAR = 10, or a factor of 1.3 when PAR = 2000. Of course, the importance of this term also depends on the magnitude of sqrt(K_sl * X

/ VPD). It is difficult to visualize the differences just from the parameters listed in the text. For the curious audience, it would be helpful to include a supplementary figure comparing optimised and unoptimised LRU_CAP values.
We have added the following figure to the supplement:

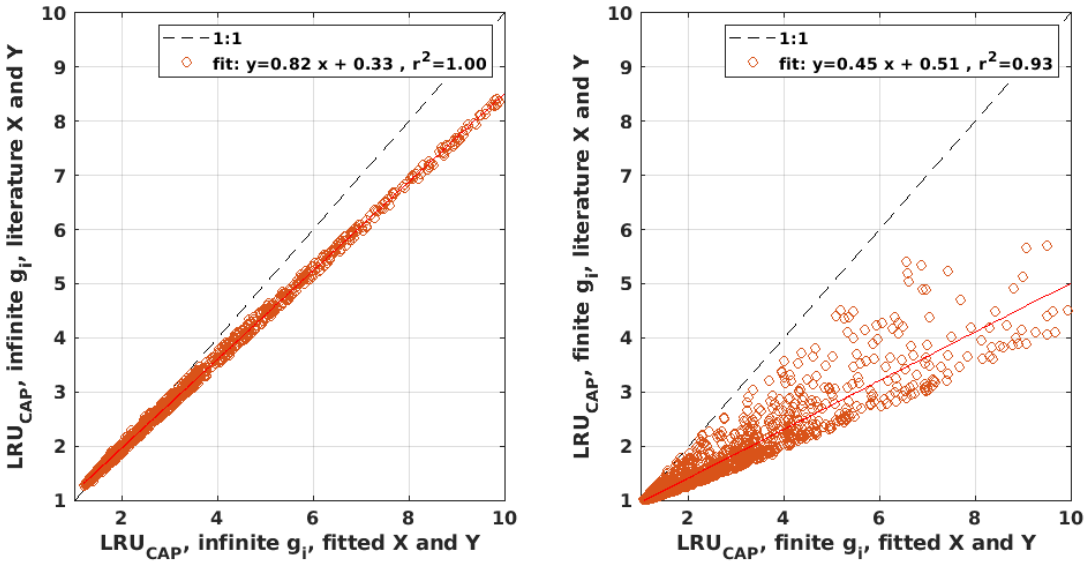

Figure B2. Scatter plots of LRU_{CAP} using the literature values against LRU_{CAP} using the optimized parameter values when assuming (a) infinite or (b) finite mesophyll conductance.

L296--297: "GPP_{COS,PAR} thus has a stronger radiation response than the other GPP estimates" - It may help to remind the audience that this is caused by lower empirical LRU estimates than the CAP estimates at high light, citing Figure B6.
We have corrected the sentence as "*$GPP_{COS,PAR}$ thus has a stronger radiation response than the other GPP estimates, due to lower empirical LRU estimate than LRU_{CAP} at high PAR (Fig. B6).*"

L312: "Because GPP_{ANN} is based purely on data, it is highly sensitive to uncertainty in the input data." - It may depend on the model architecture and what tricks are used to suppress overfitting and make the model generalisable. Suffice it to say, it is more sensitive to uncertainty in the input data than are parametric partitioning methods.
Thanks, we agree and changed the sentence in "*Because ANN fitting is purely based on the provided examples, GPP_{ANN} could be more sensitive to the uncertainty of (training) data with respect to the parametric partitioning methods.*"

L317--318: "as a different PAR relation was found in Yang et al. (2018) and it is not known if this relation holds elsewhere" - Since the LRU--PAR relationship is empirical, people can choose the equation that best fits their data, be it exponential, power law, or hyperbolic (y ~ 1/x). Therefore, it probably won't be a big problem if the empirical equation used in this study does not fit data from other sites well, so long as LRU shows a typical decreasing trend with PAR.
The reviewer is correct. We suggest to change the sentence in "*This PAR relation is site specific and different compared to the one found by Yang et al., (2018). For this reason it is not known if and how it can be used in other sites, where it is suggested to retrieve it directly from observations. The choice of empirical LRU-PAR relation at any given site is to some extent arbitrary.*"

Figure 2: "traditionally partitioned" sounds vague. I would go with "GPP partitioned using a combined nighttime-daytime method".

Corrected as suggested.

Figure 5: "All medians have been calculated using the same number of data points" - This statement seems unclear at first glance. I would rephrase it as "Medians from different methods have been calculated using the same number of data points for each month."
Corrected as *"Median differences have been calculated using the same number of data points for each method in each month."*

Figure B2: "black line" -> "black solid line"
Corrected as suggested.

Reviewer #2 Georg Wohlfahrt

Overview:
I am not one of the previous two reviewers, but was asked to review the revised version by the handling editor since one of the original reviewers apparently declined to re-review the revised version. I have quickly glanced over the reply to the reviews and my impression is that the authors have responded adequately to all reviewer comments. Regarding one comment by reviewer #1 on whether the daytime approach assumes respiration is the same during day and night, the authors are correct. While the daytime approach gets the temperature sensitivity from the nighttime data and the base respiration is based on daytime data only, the resulting relationship is applied both day and night (same for nighttime approach).
My own overall assessment of this manuscript is that this is a significant contribution to the literature adding to our understanding of the uncertainty related to partitioning NEE into its component fluxes and especially the differences between CO2 and COS-based approaches. Overall, I have not much to critique and think that the manuscript can be accepted after a few more minor changes listed below. The text is mostly well written, apart from some random switches in tenses, which may however be corrected for during copy-editing.
We thank the reviewer for the comments and suggestions to improve the manuscript.

Detailed comments:
l. 6: also the CO2 flux partitioning is applied to the same boreal forest site – suggest to introduce the site before explaining the 4 approaches
Corrected as suggested: *"In this study, we evaluate four different methods for estimating photosynthesis at a boreal forest at the ecosystem scale, of which two are based on carbon dioxide (CO_2) flux measurements and two on carbonyl sulfide (COS) flux measurements."*

l. 14: since the true GPP is unknown this would need to be acknowledge in this sentence, e.g. "… underestimation of midday GPP relative to xy."
Added *"...relative to other GPP methods."*

l. 36: for clarity I would write "… (as in nighttime method) …"
Corrected as suggested.

l. 38: the Lasslop et al. approach is typically referred to as the daytime method/approach for flux partitioning
Corrected as suggested.

l. 42: this was first quantitatively assessed by Wohlfahrt et al. (2005; 10.1016/j.agrformet.2005.02.001)

Reference added.

l. 116-121: this approach amounts to the nighttime flux partitioning logic
Added clarification "*...where R was estimated as in the nighttime method...*"

l. 120: arithmetic mean of some weighting involved?
Corrected as "*arithmetic mean between*"

l. 125: "forces GPP to zero when Ta < 0°C" – the decrease of f(Ta) already starts at positive temperatures (e.g. 5°C) and at Ta = -2 Eq. 4 yields a value of 0.5 – either Eq. 4 is wrong or the sentence in l. 125 should be reformulate to precisely convey how Eq. 4 responds to temperature
We have clarified the sentence as "*...$f(T_a)$ is an instantaneous temperature response that brings GPP gradually towards zero at freezing temperatures...*"

l. 129-130: that means you fit the parameters of Eq. 3 to the GPP that was derived from the nighttime flux partitioning (Eq. 1)? If so this should be clearly stated
Yes exactly, this is now clarified as "*The parameters of Eq. 3 were estimated from GPP partitioned with the nighttime method in Eq. 1.*"

l. 156-159: I would start here saying what the average flux was, followed by a description of the (seasonal/diurnal) variability and then continue by saying that the average soil uptake was deducted from the ecosystem-scale flux to yield the canopy net uptake
We have altered the text as suggested: "*Based on previous soil chamber measurements at Hyytiälä forest it is known that the soil COS flux was -2.7 pmol m^{-2} s^{-1} on average with a variation of only 1 pmol m^{-2} s^{-1} during the growing season and a negligible diurnal variation (Kooijmans et al., 2019; Sun et al., 2018a). The average soil flux was thus first subtracted from the quality filtered and gap-filled COS EC fluxes in order to derive the vegetation contribution to the ecosystem COS exchange.*"

l. 159: shouldn't this be "… from the canopy COS fluxes …"?
Yes, corrected as suggested.

l. 162-163: in my view the "respectively" should be placed after the concentrations, not at the end of the sentence
Corrected as suggested.

l. 226: "… may take on negative values due to …"
Corrected as suggested.

l. 228: specify how small
We added that the relative difference was 2% on average during summer months.

l. 254: also boundary layer conductance is assumed infinite in Eq. 8
Finite boundary layer conductance is now also listed as a possible improvement in the model: "*This mismatch suggests there may be scope for further model improvement, such as the inclusion of dark respiration and/or finite mesophyll and boundary layer conductances in the LRU_{CAP} model.*"

l. 286-287: I am not convinced that this will be visible in a plot of NEE as a function of PAR as on the canopy-scale this will be a gradual response as vertically in the canopy leaves/needles experience different PAR levels and thus for the same above-canopy PAR the part of the needles experiencing enough radiation may be inhibited, while others that are more shaded not – this has

been demonstrated with a multi-layer canopy model in Wohlfahrt et al. (2005; 10.1016/j.agrformet.2005.02.001)
The reviewer is correct, and we have added a sentence about this to the revised MS: "*While it is possible that less radiated needles experience less inhibition than well radiated, that cancel out at the ecosystem scale (Wohlfahrt et al., 2005), this test provides some insight to the problem.*"

l. 319: in addition there is a scaling issue here – LRU depending in a non-linear fashion on PAR means that vertically in the plant canopy where needles experience PAR that may be different (lower) compared to the above-canopy measurements and thus vertically needles operate at different LRUs; the discussion in l. 339-346 misses this crucial point
We have added a mention about the scaling issue in line 319: "*...it does not take into account the different light conditions inside the canopy, stomatal regulation during drought, or the effects of non-stomatal limitations on photosynthesis.*" and a sentence about the possible different LRUs at the end of the section: "*However, it is also possible that LRU varies throughout the canopy due to different light conditions.*"

l. 326: "… soil COS exchange …"
Corrected as suggested.

section 4 – conclusions: this section is mostly (l. 348-361) a summary rather than providing conclusions and would profit from expanding the outward looking part (what do these results mean for carbon cycle science?) at the expense of the repetition of the major results (which could be further condensed)

We think that a short summary of the results are still useful in the conclusions, but we shortened the summary and added a new paragraph on the meaning and future steps.

"*Daily GPP_{ANN} and GPP_{NLR} did not differ significantly, and differences were also small on sub-daily and seasonal time scales. $GPP_{COS,PAR}$ was higher than GPP_{NLR} on all time scales studied, including the estimate of three-month cumulative GPP during the peak growing season. In contrast, $GPP_{COS,CAP}$, a new method based on stomatal optimization theory, gave better agreement with GPP_{NLR} on all time scales, and was also less scattered than $GPP_{COS,PAR}$ on a 30-min time scale.*

The LRU_{CAP} function provides a new theoretical underpinning for COS-based GPP estimates that can be used at other measurement sites, potentially without requiring additional branch chamber measurements. LRU_{CAP} represents a significant improvement on previous LRU functions based on site-specific empirical regressions. However, LRU_{CAP} overestimated LRU at high radiation, when compared to LRU observations at the top of the canopy, leading to a lower midday $GPP_{COS,CAP}$, especially in summer. This discrepancy may result from the assumption of infinite mesophyll conductance, or the absence of dark respiration, in the underlying stomatal optimization model. LRU_{CAP} would benefit from further testing at other measurement sites with COS and CO_2 branch flux measurements, including measurements inside the canopy for better canopy-integrated LRU estimates.

Although COS flux measurements are noisier, more expensive and more difficult than those of CO_2, they provide an opportunity for better process-based understanding of photosynthesis, in comparison with more traditional CO_2-based estimates of GPP. In addition to COS, other proxies such as solar induced fluorescence and isotopic flux measurements should be tested simultaneously to properly investigate their deficiencies and advantages in estimating GPP and processes underlying photosynthesis.

The establishment of large long-term ecosystem Research Infrastructures (e.g. ICOS, NEON, TERN, see Papale 2020) – involving sites equipped with eddy covariance systems that could potentially also host COS, SIF and isotope sensors – together with the planned launch of the FLEX satellite in 2025 (https://earth.esa.int/eogateway/missions/flex) that will provide global vegetation

fluorescence measurements, open up a new phase in monitoring and understanding plant photosynthesis. Our results also underline the important role of small-scale ecophysiological measurements and models in underpinning these larger-scale initiatives."